

- (iii) यदि प्रति छात्र व्यय 20 रुपये है, तो वेतन का औसत पूर्वानुमानित मूल्य (mean forecast value) ज्ञात कीजिये। साथ ही, दिए गए व्यय आंकड़े के लिए वेतन के वास्तविक पूर्वानुमानित औसत मूल्यों के लिए 95 प्रतिशत विश्वास अंतराल स्थापित कीजिये।



[This question paper contains 24 printed pages.]

**Your Roll No.....**

**Sr. No. of Question Paper : 6124**

**J**

Unique Paper Code : 2272102403

Name of the Paper : Introductory Econometrics

Name of the Course : NEP

Semester : IV

Duration : 3 HourS

Maximum Marks : 90

समय : 3 घण्टें

पूर्णांक : 90

**Instructions for Candidates**

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. All questions carry equal marks.
3. Use of a simple non-programmable calculator is allowed. Statistical tables are attached for your reference.
4. All intermediate calculations should be rounded off to 4 decimal places.

5. The values provided in statistical tables should not be rounded off.
6. All relevant tables are attached to the paper All final calculations should be rounded off to two decimal places.
7. Question 1 is compulsory, and attempt any **four** questions from the remaining **six** questions.
8. Answers may be written either in English or Hindi; but the same medium should be used throughout the paper.

#### छात्रों के लिए निर्देश

1. इस प्रश्न-पत्र के मिलते ही ऊपर दिए गए निर्धारित स्थान पर अपना अनुक्रमांक लिखिए।
2. सभी प्रश्नों के अंक समान हैं।
3. सरल गैर-प्रोग्रामेबल कैलकुलेटर का उपयोग करने की अनुमति है। आपके संदर्भ के लिए सांख्यिकीय तालिकाएँ संलग्न हैं।
4. सभी मध्यवर्ती गणनाएँ 4 दशमलव स्थानों तक पूर्णांकित की जानी चाहिए।

$Y_i = \beta_1 + \beta_2 X_i + u_i$ , जहाँ  $Y$  का अर्थ है शिक्षक का वेतन (रुपये में) और  $X$  का अर्थ है प्रति छात्र व्यय (रुपये में)। निम्नलिखित डेटा उपलब्ध है:-

$$\bar{Y} = 3.09 \quad \bar{X} = 9 \quad \sum y_i x_i = -15.1$$

$$\sum x_i^2 = 116 \quad \sum y_i^2 = 2.27 \quad n = 7$$

(जहाँ  $X$  और  $y$  को विचलन रूप (deviation form) में दिया गया है।)

- (i) मापदंडों और उनकी मानक त्रुटियों के अनुमान प्राप्त कीजिये।
- (ii) ढलान गुणांक के लिए 95 प्रतिशत विश्वास अंतराल (confidence interval) स्थापित कीजिये। क्या आप इस परिकल्पना को अस्वीकार करेंगे कि वास्तविक ढलान गुणांक 0 है?

(i) मॉडल के अवरोधन गुणांक (intercept coefficient) की व्याख्या कीजिये। आय का महत्वपूर्ण या सीमा (threshold) स्तर और उपभोग का संतुष्टि स्तर क्या है?

(ii) यदि उपरोक्त प्रतिगमन में प्राप्त त्रुटियाँ उच्च-क्रम स्वप्रतिगमन (higher-order autoregressive) प्रक्रिया का अनुसरण करती हैं, तो प्रथम-क्रम क्रमिक सहसंबंध की जाँच के लिए एक परीक्षण का वर्णन कीजिये। परीक्षण के चरणों को विस्तार से बताइएँ।

(iii) प्रतिगमन मॉडल  $Y_t = \beta_1 + \beta_2 \frac{1}{X_t} + u_t$ , में, यदि  $\rho$  ज्ञात हो, तो क्रमिक सहसंबंध की समस्या का संभावित समाधान क्या होगा?

7. यह पता लगाने के लिए कि क्या सरकारी स्कूलों में शिक्षकों के वेतन और प्रति छात्र व्यय के बीच कोई संबंध है, निम्नलिखित मॉडल का सुझाव दिया गया:

5. सांख्यिकीय तालिकाओं में दिए गए मानों को पूर्णांकित नहीं किया जाना चाहिए।
6. सभी प्रासंगिक तालिकाएँ पेपर के साथ संलग्न हैं। सभी अंतिम गणनाएँ दो दशमलव स्थानों तक पूर्णांकित की जानी चाहिए।
7. प्रश्न 1 अनिवार्य है, और शेष छह प्रश्नों में से कोई भी चार प्रश्न हल करें।
8. इस प्रश्न-पत्र का उत्तर अंग्रेजी या हिंदी किसी एक भाषा में दीजिए, लेकिन सभी उत्तरों का माध्यम एक ही होना चाहिए।

1. State whether the following statements are True or False. Give reasons/justifications for your answer.

- (i) In simple linear regression models,  $r^2$  value is invariant to changes in the unit of measurement.
- (ii) In a multiple regression model  $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ , testing a joint restriction  $H_0: \beta_2 = \beta_3 = 0$  is the same as testing for  $H_0: \beta_2 = 0$  and  $H_0: \beta_3 = 0$ .

- (iii) In the presence of heteroscedasticity, the usual OLS method always overestimates the standard errors of estimators.
- (iv) In a bivariate regression, the F-statistic is equal to the t-statistic for the slope coefficient.
- (v) Consider the following multiple linear regression model :  $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ . Assume that the 95% confidence interval for  $\beta_2$  is  $[-0.254, 0.878]$ . Therefore,  $\beta_2$  is statistically different from 1 at a 5% significance level.
- (vi) A sample correlation coefficient of 0.95 between two independent variables, both included in the regression model, causes OLS estimators to be biased.
2. A researcher obtained the following ordinary least squares (OLS) estimates for a firm's stock price using monthly data from 2000 M1 to 2009 M12 (all

तो प्रतिगमन II में विनिर्देश त्रुटि की संभावना की जाँच करने के लिए रैमसे रीसेट परीक्षण कीजिये। 10% महत्व (significance) स्तर का उपयोग कीजिये। अपने निष्कर्ष को स्पष्ट रूप से बताइए।

- (iii) मान लीजिए कि औसत घरेलू आय (I) को सही तरीके से नहीं मापा गया है। इसके क्या परिणाम होंगे? अब, यदि सकल बिक्री मात्रा (Y) को गलत तरीके से मापा जाता है, तो इसके क्या परिणाम होंगे?

6. ABC राज्य के लोगों के लिए निम्नलिखित एंगेल वक्र (Engel curve) का अनुमान लगाया गया है, जहाँ Y किसी वस्तु पर व्यय को दर्शाता है और X कुल आय को दर्शाता है। इस संबंध का आकलन वर्ष 1990 से 2022 तक लगाया गया है।

$$\hat{Y}_t = 2.8 - 6 \frac{1}{X_t}$$

$$se = (1.1) \quad (0.02)$$

$$R^2 = 0.8248 \quad d = 0.7243$$



जहाँ  $Y$ =सकल बिक्री मात्रा,  $N$ =आस-पास के प्रतिस्पर्धी रेस्तरां की संख्या,  $P$ =आस-पास की जनसंख्या, और  $I$ =आस-पास की औसत घरेलू आय है।

- (i) मान लीजिए कि मॉडल में एक और चर  $A$  को जोड़कर पुनः आकलित (re-estimated) किया जाता है, जहाँ  $A$  = रेस्तरां के पते के अंतिम दो अंक हैं। नीचे दिए गए संशोधित प्रतिगमन पर विचार कीजिये :

$$\text{Regression II: } \hat{Y}_i = 98125 - 8975 N_i + 0.360 P_i + 1.301 I_i + 58.07 A_i$$

$$\text{se} = (265) \quad (2082) \quad (0.074) \quad (0.550) \quad (95.21)$$

$$n = 33 \quad R^2 = 0.623 \quad \bar{R}^2 = 0.569$$

क्या आपको ऊपर दिए गए प्रतिगमन II में विनिर्देश त्रुटि (specification error) का संदेह है? ऐसी त्रुटि के क्या परिणाम हो सकते हैं?

- (ii) यदि शोधकर्ता अतिरिक्त प्रतिगमन के रूप में  $\hat{Y}_i$  के वर्गों और घनों के साथ प्रतिगमन II को फिर से चलाने (rerun) का निर्णय लेता है और  $R^2$  को 0.785 के रूप में प्राप्त करता है,

variables in logarithms). The following regression was obtained:

$$\ln s_t = 0.87 - 0.54 \ln p_t + 0.65 \ln y_t + 0.34 \ln r_t - 0.32 \ln m_t$$

$$\text{se} = (0.24) \quad (0.30) \quad (0.12) \quad (0.24) \quad (1.07)$$

$$R^2 = 0.34 \quad \text{RSS} = 1.24 \quad F_{4,115} = 14.91$$

$s_t$  = stock price,  $p_t$  = profits,  $y_t$  = output in the economy,  $r_t$  = expenditure on research and development, and  $m_t$  = expenditure on marketing. Figures in parentheses are standard errors, and RSS is the Residual Sum of Squares.

- (i) Interpret the partial regression coefficients of the above regression and test their joint significance at  $\alpha = 5\%$ .
- (ii) Test the hypothesis that the elasticity of stock price with respect to output is unit elastic versus it is inelastic. Use  $\alpha = 1\%$ .

- (iii) Describe a test to show no significant difference between the coefficients of expenditure on research & development and expenditure on marketing. (6,6,6)

3. (a) An investigator estimated the following demand function for cars in relation to the price of cars ( $X_2$ ) and consumer income ( $X_3$ ) using quarterly data from 2000 to 2019 as

$$\hat{Y}_i = 145.37 - 2.7975X_{2i} - 0.3191X_{3i}$$

$$se = (120.06) \quad (0.8122) \quad (0.4003)$$

$$R^2 = 0.9778$$

- (i) Are the partial regression coefficients statistically significant? Do you think there could be a problem with the above regression?

- (iii) मान लीजिए कि शोधकर्ता यह परिकल्पना करता है कि मूल मॉडल में हेटेरोस्केडैस्टिसिटी के कारण समस्या आयेगी और व्हाइट टेस्ट (वर्ग और क्रॉस उत्पादों के साथ) आयोजित करता है। परीक्षण सांख्यिकी का मान 23.789 निकलता है। क्या आपके निष्कर्ष में शोधकर्ता का दावा सही है? मान लीजिए कि यह माना जाता है कि  $\sigma_i^2 = \sigma^2 \sqrt{(\text{exper}_i)}$  है, तो हेटेरोस्केडैस्टिसिटी की समस्या को कैसे ठीक किया जा सकता है?

5. शहर में सबसे अच्छे पारिवारिक रेस्तराँ का निर्धारण करने के लिए, उस रेस्तराँ की अवस्थिति के विभिन्न वर्णनकर्ताओं के फलन के रूप में विभिन्न रेस्तराँ की सकल बिक्री मात्रा को समझाने के लिए एक प्रतिगमन मॉडल का उपयोग किया जाता है। वह निम्नलिखित प्रतिगमन (मूल) पर विचार करता है:

**Regression I:**  $\hat{Y}_i = 102192 - 9075 N_i + 0.3547 P_i + 1.288 I_i$

$$se = (389) \quad (2053) \quad (0.0727) \quad (0.543)$$

$$n = 33 \quad R^2 = 0.618 \quad \bar{R}^2 = 0.579$$

(i) क्या पुरुष और महिला कर्मचारियों के वेतन में सांख्यिकीय रूप से अलग-अलग है? यदि कोई अन्य शोधकर्ता एक डमी चर को परिभाषित करता है, जो महिलाओं के लिए 1 और अन्यथा 0 का मान लेता है, तो प्रतिगमन गुणांक के अनुमान किस प्रकार बदल जायेंगे?

(ii) आप उपरोक्त प्रतिगमन में निम्नलिखित को अतिरिक्त रूप से शामिल करने के लिए मॉडल को किस प्रकार पुनः निर्दिष्ट (respecify) करेंगे:

(क) मजदूरी पर शिक्षा का लिंग-संवेदनशील प्रभाव। आप इस संवेदनशीलता के सांख्यिकीय महत्व का परीक्षण किस प्रकार करेंगे?

(ख) मान लीजिये किसी की आयु में 50 से 51 वर्ष की वृद्धि होने की तुलना में आयु में 25 से 26 वर्ष की वृद्धि होने का मजदूरी पर अधिक प्रभाव पड़ेगा? आप इस गुणांक के लिए किस संकेत (sign) की अपेक्षा करते हैं?

(ii) What are the consequences of the problem detected in part i of the question above?

(b) Show that the coefficient of determination,  $R^2$ , can also be obtained as the squared correlation between actual Y values and the Y values estimated from the regression model, where Y is the dependent variable. (6,6,6)

4. The wage rate per month (in '000 Rs.) for 50 employees (25 males and 25 females) was regressed on education (in years), experience (in years), Age (in years), and a dummy variable which takes the value 1 for males and 0 otherwise.

$$\widehat{Wage}_i = 648 + 132Educ_i + 38.0Exper_i - 5.83Age_i + 488D_i$$

$$se = \quad (383) \quad (31.7) \quad (13.0) \quad (7.69) \quad (147)$$

$$R^2 = 0.457$$

(i) Do wages of male and female employees differ statistically? How will the estimates of the regression coefficients change if another researcher defines a dummy variable, which takes the value 1 for females and 0 otherwise?

(ii) How would you respecify the model to additionally include the following in the regression above :

(a) The gender-sensitive impact of education on wages. How will you test the statistical significance of this sensitivity?

(b) Belief that an increase in age from 25 to 26 years will have more impact on wages than if the age increased from 50 to 51? What sign do you expect for this coefficient?

(i) क्या आंशिक प्रतिगमन गुणांक सांख्यिकीय रूप से महत्वपूर्ण हैं? क्या आपको लगता है कि उपरोक्त प्रतिगमन में कोई समस्या हो सकती है?

(ii) उपरोक्त प्रश्न के भाग I में पाई गई समस्या के परिणाम (consequences) क्या हैं?

(ख) निर्धारण गुणांक  $R^2$  को वास्तविक  $Y$  मानों और प्रतिगमन मॉडल से अनुमानित  $Y$  मानों के बीच वर्ग सहसंबंध (squared correlation) के रूप में भी प्राप्त किया जा सकता है, जहाँ  $Y$  आश्रित चर है, दर्शाए।

4. 50 कर्मचारियों (25 पुरुष और 25 महिलाएं) के लिए प्रति माह मजदूरी दर ('000 रु. में) को शिक्षा (वर्षों में), अनुभव (वर्षों में), आयु (वर्षों में) और एक डमी चर के आधार पर प्रतिगमन किया गया, जो पुरुषों के लिए मान 1 और अन्यथा 0 लेता है।

$$\widehat{Wage}_i = 648 + 132Educ_i + 38.0Exper_i - 5.83Age_i + 488D_i$$

$$se = \begin{matrix} (383) & (31.7) & (13.0) & (7.69) & (147) \end{matrix}$$

$$R^2 = 0.457$$



- (i) उपरोक्त प्रतिगमन के आंशिक प्रतिगमन गुणांकों की व्याख्या कीजिये और  $\alpha = 5\%$  पर उनके संयुक्त महत्व का परीक्षण कीजिये।
- (ii) इस परिकल्पना का परीक्षण कीजिये कि उत्पादन के संबंध में स्टॉक मूल्य की लोच इकाई लोचदार (unit elastic) बनाम अलोचदार (inelastic) है।  $\alpha = 1\%$  का उपयोग कीजिये।
- (iii) अनुसंधान और विकास पर व्यय और विपणन पर व्यय के गुणांकों के बीच कोई महत्वपूर्ण अंतर नहीं है यह दर्शाने के लिए एक परीक्षण का वर्णन कीजिये।
3. (क) किसी अन्वेषक ने वर्ष 2000 से 2019 तक के तिमाही आंकड़ों का उपयोग करते हुए कारों की कीमत ( $X_2$ ) और उपभोक्ता आय ( $X_3$ ) के संबंध में कारों के लिए निम्नलिखित मांग फलन का अनुमान लगाया :

$$\hat{Y}_i = 145.37 - 2.7975X_{2i} - 0.3191X_{3i}$$

$$se = (120.06) \quad (0.8122) \quad (0.4003)$$

$$R^2 = 0.9778$$

- (iii) Suppose the researcher hypothesizes that the original model is bound to suffer from heteroscedasticity and conducts White's Test (with squares and cross products). The value of the test statistic turns out to be 23.789. Would you conclude that the researcher's claim is correct?

Suppose it is postulated that of  $\sigma_i^2 = \sigma^2 \sqrt{(\text{exper}_i)}$ , how is the problem of heteroscedasticity corrected? (6,6,6)

5. To decide the best family restaurants in the city, a regression model is used to explain the gross sales volume of various restaurants as a function of different descriptors of the location of that restaurant. He considers the following regression (original) :

**Regression I:**  $\hat{Y}_i = 102192 - 9075 N_i + 0.3547 P_i + 1.288 I_i$

se = (389)      (2053)      (0.0727)      (0.543)

n = 33       $R^2 = 0.618$        $\bar{R}^2 = 0.579$

where  $Y$  = gross sales volume,  $N$  = the number of competitive restaurants nearby,  $P$  = the population nearby, and  $I$  = the average household income nearby.

- (i) Suppose the model is re-estimated by adding another variable  $A$ , where  $A$  = the last two digits of the restaurant's address. Consider the modified regression given below :

$$\text{Regression II: } \hat{Y}_i = 98125 - 8975 N_i + 0.360 P_i + 1.301 I_i + 58.07 A_i$$

$$\text{se} = (265) \quad (2082) \quad (0.074) \quad (0.550) \quad (95.21)$$

$$n = 33 \quad R^2 = 0.623 \quad \bar{R}^2 = 0.569$$

Do you suspect a specification error in Regression II above? What could be the consequences of such an error?

- (ii) Carry out the Ramsey RESET test to check for the likelihood of specification error in Regression II if the researcher decides to rerun Regression II with squares and cubes of  $\hat{Y}_i$  as additional regressors and obtains  $R^2$  as 0.785. Use a 10% level of significance. State your conclusion clearly.

- (vi) दो स्वतंत्र चरों, दोनों प्रतिगमन मॉडल में शामिल, के बीच 0.95 का एक नमूना सहसंबंध गुणांक OLS अनुमानकों को पक्षपाती बनाता है।

2. एक शोधकर्ता ने 2000 M1 से 2009 M12 तक के मासिक डेटा (सभी चर लघुगणक में) का उपयोग करके एक फर्म के स्टॉक मूल्य के लिए निम्नलिखित साधारण न्यूनतम वर्ग (OLS) अनुमान प्राप्त किए। निम्नलिखित प्रतिगमन प्राप्त किया गया :

$$\ln s_t = 0.87 - 0.54 \ln p_t + 0.65 \ln y_t + 0.34 \ln r_t - 0.32 \ln m_t$$

$$\text{se} = (0.24) \quad (0.30) \quad (0.12) \quad (0.24) \quad (1.07)$$

$$R^2 = 0.34 \quad \text{RSS} = 1.24 \quad F_{4,115} = 14.91$$

$s_t$  = स्टॉक प्राइस,  $p_t$  = लाभ,  $y_t$  = अर्थव्यवस्था में उत्पादन,  $r_t$  = अनुसंधान एवं विकास पर व्यय, और  $m_t$  = विपणन पर व्यय है। कोष्ठक में दिए गए आंकड़े मानक त्रुटियाँ हैं, तथा RSS वर्गों का अवशिष्ट योग है।

(ii) एकाधिक प्रतिगमन मॉडल  $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$  में, संयुक्त प्रतिबंध (restriction)  $H_0: \beta_2 = \beta_3 = 0$  का परीक्षण करना  $H_0: \beta_2 = 0$  और  $H_0: \beta_3 = 0$  का परीक्षण करने के समान है।

(iii) हेटेरोस्केडैस्टिसिटी (heteroscedasticity) की उपस्थिति में, सामान्य OLS विधि हमेशा अनुमानकों की मानक त्रुटियों को अधिक आंकती है।

(iv) किसी द्विचर प्रतिगमन (bivariate regression) में, ढलान गुणांक के लिए F-सांख्यिकी t-सांख्यिकी के बराबर होती है।

(v) निम्नलिखित एकाधिक रैखिक प्रतिगमन मॉडल पर विचार कीजिये:  
 $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$  मान लीजिये कि  $\beta_2$  के लिए 95% विश्वास अंतराल (confidence interval)  $[-0.254, 0.878]$  है। इसलिए,  $\beta_2$  5% महत्व स्तर पर 1 से सांख्यिकीय रूप से भिन्न है।

(iii) Suppose average household income (I) is not measured correctly. What are the consequences? Now, if the gross sales volume (Y) is incorrectly measured, then what are the consequences?

(6,6,6)

6. Following Engel curve was estimated for people of the ABC state, where Y denotes expenditure on a commodity and X the total income. The relationship was estimated from 1990-2022.

$$\hat{Y}_t = 2.8 - 6 \frac{1}{X_t}$$

$$se = (1.1) \quad (0.02)$$

$$R^2 = 0.8248 \quad d = 0.7243$$

(i) Interpret the intercept coefficient of the model. What is the critical or *threshold* level of income and the satiety level of consumption?

(ii) If the errors obtained in the above regression follow a higher-order autoregressive process, then describe a test to check for first-order serial correlation. Give the steps of the test in detail.



- (iii) In a regression model  $Y_t = \beta_1 + \beta_2 \frac{1}{x_t} + u_t$ , what would be the possible remedy to the problem of serial correlation if  $\rho$  is known? (6,6,6)

7. To find out if there is any relationship between teachers' salary and per-student expenditure in public schools, the following model was suggested :

$Y_i = \beta_1 + \beta_2 X_i + u_i$ , where Y stands for teacher's salary (in Rs.) and X stands for per student expenditure (in Rs.). The following data is available

$$\bar{Y} = 3.09 \quad \bar{X} = 9 \quad \sum y_i x_i = -15.1$$

$$\sum x_i^2 = 116 \quad \sum y_i^2 = 2.27 \quad n = 7$$

(where x and y are given in deviation form)

- (i) Obtain the estimates of the parameters and their standard errors.

- (ii) Establish a 95 percent confidence interval for the slope coefficient. Would you reject the hypothesis that the true slope coefficient is 0?

- (iii) Obtain the mean forecast value of salary if per-student spending is Rs. 20. Also, establish a 95 percent confidence interval for the true predicted mean values of salary for the given spending figure.

1. निम्नलिखित कथन सही हैं या गलत। अपने उत्तर के पक्ष में कारण/औचित्य दीजिए।

- (i) सरल रेखीय प्रतिगमन (simple linear regression) मॉडल में,  $r^2$  मान माप की इकाई में बदलावों के प्रति अपरिवर्तनीय (invariant) है।